



Jiang, D. (2016). A Critique of the Journal Article 'Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People's Republic of China' by Charles Teddlie and Shujie Liu (2008). *International Journal of Education*, 8(1), 94-106.
<https://doi.org/10.5296/ije.v8i1.8793>

Publisher's PDF, also known as Version of record

Link to published version (if available):
[10.5296/ije.v8i1.8793](https://doi.org/10.5296/ije.v8i1.8793)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Macrothink Institute at <http://dx.doi.org/10.5296/ije.v8i1.8793>.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

A Critique of the Journal Article ‘Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People’s Republic of China’ by Charles Teddlie and Shujie Liu (2008)

Dini Jiang^{1,*}

¹Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, United Kingdom

*Correspondence: Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, United Kingdom. Tel: 44-7928-530-914 E-mail: dini.jiang@bristol.ac.uk

Received: December 28, 2015 Accepted: January 16, 2016 Published: March 22, 2016

doi:10.5296/ije.v8i1.8793 URL: <http://dx.doi.org/10.5296/ije.v8i1.8793>

Abstract

This paper will evaluate the journal article ‘Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People’s Republic of China’, co-authored by Charles Teddlie and Shujie Liu and published in the journal ‘School Effectiveness and School Improvement’ in 2008. It presents the philosophical classification of the research text, with a particular focus on the underlying view of post-positivism. It outlines two different types of evaluation criteria and, based on the criteria, identifies both the strengths and weaknesses of the research text respectively. It discusses the improvements that could be made for improving the research and concludes with reflections on how the philosophical approach could impact on the coherence between research design and research methodology as well as the recognition of the advantages and disadvantages of combining quantitative and qualitative methods in a piece of mix-methods research.

Keywords: Teacher effectiveness, China, mixed-methods research, post-positivism, critique

1. Introduction

This paper will evaluate the journal article ‘Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People’s Republic of China’, co-authored by Charles Teddlie and Shujie Liu and published in the journal *School Effectiveness and School Improvement* in 2008. This text has been selected based on the following criteria: 1) The study is a good example of how to employ a mixed-methods approach to research teacher effectiveness; 2) It has contributed greatly to both the literature on teacher effectiveness research (TER) in the Chinese context, where few empirical studies have been conducted (Peng et al., 2014; Thomas et al., 2012; Thomas & Peng 2011; Sun, 2011; Teddlie & Liu, 2008; Sun, 2007; Teddlie et al., 2006), and to the research evidence that international TER researchers could draw on for further research (Thomas et al., 2015; Thomas et al., 2012; Reynolds, 2006; Teddlie et al., 2006).

Specifically, the study investigated teacher effectiveness within differentially effective (more effective and less effective) primary schools in China. The aim of the study has been to ‘advance our understanding of educational effectiveness processes’ (Teddlie & Liu, 2008, p. 387) by exploring the differences of effective schooling in two different Chinese community types (urban and rural). The paper uses a contextually sensitive effectiveness research design and a parallel mixed-methods methodology (QUAN + qual, see Johnson & Onwuegbuzie, 2004), with an emphasis on the quantitative data, is employed to identify differences in teacher effectiveness for crossing 2 levels of effectiveness status and 2 levels of community types. The instruments of the Classroom Snapshot (CS) and the Louisiana Components of Effective Teaching (LCET) Summary Form have been used to collect the quantitative data and informal notes have been taken in classrooms, together with school-level observations and other documents, to gather the qualitative data (Teddlie & Liu, 2008). The research findings are 1) there has been positive evidence for Hypothesis #1 that ‘Primary schools designated as more effective will have teachers who exhibit more effective teaching practices than schools designated as less effective’ (Teddlie & Liu, 2008, p. 391); 2) there has also been positive evidence for Hypothesis #2 that ‘Primary schools from urban areas will have teachers who exhibit more effective teaching practices than schools from rural areas’ (Teddlie & Liu, 2008, p. 392); and 3) the effective teaching practices which are ‘indigenous to China’ have also been identified as follows:

A fundamental reliance on whole class interactive teaching, an emphasis on discipline which is multilayered and essential to educational effectiveness in China, the role of the *banzhuren*, difficulties in promoting a positive learning environment for all students, and widespread teaching to the test (Teddlie & Liu, 2008, p. 401).

This paper aims to 1) highlight the importance of employing the most appropriate philosophical and methodological approaches in planning and designing research; and 2) recognise both the advantages and disadvantages of combining quantitative and qualitative methods for a mixed-methods study.

Overall, this paper will be presented in four sections. Section 2 presents the philosophical classification of the researcher text, with a particular focus on the underlying view of

post-positivism. Section 3 outlines two different types of evaluation criteria and, based on the criteria, identifies both the strengths and weaknesses of the research text respectively. Section 4 discusses the improvements that could be made for improving the research and section 5 concludes with reflections on how the epistemological approach could impact on the coherence between research design and research methodology as well as the recognition of the advantages and disadvantages of combining quantitative and qualitative methods in a piece of mix-methods research.

2. Philosophical classification of the research text: Post-positivism

Generally speaking, it is argued that a post-positivistic philosophical approach has been employed in this text. Before commencing, it is, however, important to set out the underlying view of post-positivism as a basis for classifying the philosophical assumption of the research text.

First, post-positivists accept that there is a real reality, but only a certain level of objectivity exists, rather than absolute objectivity. Based on an ontological perspective, Bohr has questioned what the absolute objectivity is, urging us to embrace the ambiguity of scientific knowledge (Crotty, 1998). Popper proposed the principle of falsification to argue that ‘scientific truth turns out to be, not something shown to be true, but simply something that scientists have so far been unable to prove false’ (Crotty, 1998, p. 33). Guba and Lincoln have described the ontological assumption of post-positivism as ‘real reality but only imperfectly and probabilistically apprehendible’ (Guba & Lincoln, 2005, p. 193), thereby providing us with deeper insights into what a certain level of objectivity means.

Second, instead of seeking certainty, post-positivists emphasise probability and tentativeness. According to Heisenberg’s epistemological argument, the uncertainty principle, it is very difficult for scientists to determine the nature of knowledge with scientific objectivity solely based on what we human beings know (Crotty, 1998). The contradiction lies in whether scientists are just passively noting laws or universality that are found in nature or they are actively constructing scientific knowledge in a way that is tentative. Would research findings be absolutely true or just probably be true if based on a post-positivistic assumption?

Third, post-positivism allows a certain degree of interpretation. As argued by Kuhn, scientific endeavour is a very human affair and research is led by a paradigm (Kuhn, 1970). Crotty (1978) has further argued, based on a historical and sociological stance, that research is a social process, rather than a logical one. Importantly, a social process involves the interpretations of human beings, and researchers may want to employ qualitative methods to better explore their research questions inductively (Guba & Lincoln, 2005).

To conclude, perhaps Crotty has given us a good definition of post-positivism in his work of ‘the Foundations of Social Research’:

...Scientists ‘from within’...challenged its claims to objectivity, precision and certitude...this is a less arrogant form of positivism. It is one that talks of probability

rather than certainty, claims a certain level of objectivity rather than absolute objectivity, and seeks to approximate the truth rather than aspiring to grasp it in its totality or essence (Crotty, 1998, p. 29).

Clearly, Crotty (1998) has highlighted the importance of acknowledging the relativeness of absolute objectivity whilst understanding social reality. Post-positivism recognises the limitations of positivism by asserting that knowledge claims need to be justified based on a social dimension of historical and cultural locatedness when positivists pursue lawful and universal rules (Usher, 1996, also see Philips 2004).

It is based on this underlying view of post-positivism that I would argue that a post-positivistic approach has been employed in this study. The reasons are as follows: 1) the research hypotheses and questions have been proposed to explore the probability of the differences in effective schooling between urban areas of China and rural areas of China. It is a tentative study of examining the relevance of TER factors in mainland China; 2) the authors have reflected that they have just scratched the surface of TER research in the Chinese context and that the field of the research calls for more qualitative work of investigating differentiated teacher effectiveness (see for example Campbell, 2003).

Nevertheless, one would probably argue, if only based on the aim of the study (to advance our understanding of educational effectiveness processes), that the authors could have employed an interpretive philosophical approach. From an interpretive philosophical assumption, all human action is meaningful and it is crucial to understand the meanings that construct and are constructed by interactive human behaviours (Usher, 1996). Just as argued by Crotty, the interpretivist approach looks for 'culturally derived and historically situated interpretations of the social life world' (Crotty, 1998, p. 67). That means, when we explain the social world, we have to understand it and make sense of it within a specific cultural or social context. In this study, the authors have aimed to advance our understanding of educational effectiveness processes in the Chinese context, seeking to understand differentiated teacher effectiveness factors in two different community types. Therefore it could have been argued that an interpretive philosophical approach has been employed.

3. Evaluation criteria of the research text and its strengths and weaknesses

Overall, the research text will be evaluated on the basis of the Research Excellence Framework (REF) criteria (REF, 2012) and Punch's evaluative criteria (Punch, 2009). The REF criteria will be selected to evaluate the research text with regard to its originality, significance and rigour, whereas Punch's evaluative criteria will be used to illustrate more specifically the research processes. The former will provide broader evaluation criteria for evaluating the research in general, whilst the latter will focus more specifically on the set up of the research, empirical procedures, the quality of the data, the findings and conclusions as well as the presentation of the research. Table 1 below will present the REF criteria and table 2 will outline the Punch's evaluation criteria. My reasons against each criterion will also be presented. The strengths and weaknesses will be identified.

3.1 REF evaluation criteria and relevant strengths and weaknesses

In this section, the REF evaluation criteria will be presented with the reasons given by myself in Table 1. The relevant strengths and weaknesses will also be presented respectively.

Table 1. REF evaluation criteria

Criteria (REF, 2012)	Reasons
Originality	It is important for the reader to understand in general whether the research text is original enough to be considered as an interesting intellectual contribution.
Significance	Whether the research text is important in contributing to the theoretical and methodological developments or has important policy implications.
Rigour	The extent to which the research is rigorously planned, developed and carried out, and whether the research aims, questions, methods and findings are consistent with one another (Punch & Onacea, 2014; Punch, 2009).

3.1.1 Strengths

In general, the research study is very original in shedding a new light on TER research in an emerging context like mainland China where few empirical TER studies have been conducted. It is significant both for policy makers in China to understand TER issues in more depth and for international TER researchers to look at the differentiated teacher effectiveness outside of the traditional, western context, within which more participation of TER researchers is needed (Thomas et al., 2015; Thomas et al., 2012). Overall, regardless of the research aim, the study has been designed and conducted quite rigorously. This study is a partial replication of the school effectiveness study conducted in the U.S. by Teddlie and Stringfield (1993) and the International School Effectiveness Research Project (ISERP) (Reynolds et al., 2002). The research instrument has been tested across different contexts and therefore has had high reliability and validity. Furthermore, the researchers have been transparent in presenting how they have designed and conducted the research.

3.1.2 Weaknesses

In general, a lack of coherence would be a main critique of the research text (REF, 2012; Punch, 2009). The research aim and hypotheses have been inconsistent. The research has been set up to ‘advance our understanding of educational effectiveness processes’ (Teddlie & Liu, 2008, p. 387), which implies a more qualitative approach (Punch, 2009), but the research questions have been hypothesised and addressed in a more quantitative way.

3.2 Punch’s evaluative criteria

Similarly, Punch’s evaluative criteria will be presented in Table 2, together with the reasons

given by myself. The strengths and weaknesses will also be presented accordingly.

Table 2. Punch's evaluative criteria

Main areas (Punch, 2009)	Criteria (Punch, 2009)	Reasons
The set up of the research	Is it clear what position the research is coming from?	The epistemological assumptions would have a big impact on the research design of the study.
	Are the area and topic clearly identified?	It shows what key research literature would be built on and how the research questions would be put forward. Based on what literature gaps?
	Are the research questions appropriate?	Good research questions would be helpful for the reader to understand how the study will be planned, designed and carried out.
	Is the research appropriately set in context?	'Context matters' (Crossley, 2010; 2014). Would the research questions be relevant to the theoretical/practical concern? And how the relevant literature would be handled appropriately (Punch, 2009)
Empirical procedures	Are the design, data collection and data analysis procedures reported in sufficient detail to enable the research to be replicated?	Good empirical procedures would enable the reader to examine all the stages of the study thoroughly.
	Is the design of the study appropriate for the research questions?	This relates to the importance of research questions. As long as the research questions are put forward, the study needs to be designed purposefully to address all the questions.
	Are the data collection instruments and procedures adequate and appropriate for the research questions?	Different collection instruments and procedures would generate different data, thereby influencing substantively the quality of the data.
	Is the sample appropriate?	Again, the extent to which the findings are reliable, valid or reactive depends on how the data would be sampled appropriately (Teddle & Yu, 2007).
	Are the data analysis procedures adequate and appropriate for the research questions?	To be transparent in analysing the data is crucial. Researchers need not to manipulate the procedures for research findings serving the researchers' own subjective purposes but to analyse the data as objectively as possible.

The quality of the data	Procedures in the collection of the data	Again, different collection procedures would influence the quality of the data.
	Technical aspects of the quality of the data	Reliability basically means consistency. Whether the quality of the data has consistency over time or there is internal consistency of a measuring instrument (Punch, 2009). The greater the reliability is, the closer to true scores the findings would be.
	Validity	The key question is ‘how do we know that this measuring instrument measures what we think it measures?’ (Punch, 2009, p.246). Is the data valid enough to generalise the research findings?
	Reactivity	It concerns ‘the extent to which the process of data collection changes the data’ (Punch, 2009, p. 313).
The findings and conclusions reached in the research	Have the research questions been answered?	It is the objective of the research to answer the research questions.
	How much confidence can we have in the answers put forward?	This is concerned with the generalisability of the findings and reminds us of the importance of ‘rigour’ in research.
	What can be concluded from the research on the basis of what was found?	This relates to the applicability of the findings.
Presentation		Is the research presented clearly?

3.2.1 The set up of the research

The set up of the research will be evaluated on the basis of the four aspects outlined in table 2: 1) research positioning; 2) research area and topic; 3) research questions and 4) literature review. The strengths and weaknesses will be identified respectively.

3.2.1.1 Strengths

The area and topic have been clearly identified. The study has been carried out deductively, from the topic of teacher effectiveness to research questions. In addition, the research has been set in context appropriately, with a contextually sensitive research design. The literature regarding school effectiveness and teacher effectiveness has been reviewed quite systematically both in the Chinese context and in the international context. The research questions have been put forward based on the emphasis of the context-specific effect of community type (urban and rural) at the end of the literature review.

3.2.1.2 Weaknesses

The philosophical assumption behind the research is confusing. The study has been set up to advance our understanding of educational effectiveness processes within differentially effective primary schools in mainland China, implying a more interpretivistic (qualitative) approach (Punch & Onacea, 2014), but the study has been designed and conducted with a post-positivistic approach.

Also, the research questions have not been appropriate enough to address the research aim of the study. The questions have not captured the multidimensional nature of teacher effectiveness. The effectiveness status of schools or community types could be used to conceptualise the differentiated teacher effectiveness factor in the Chinese context but perhaps would not be able to fully capture how complicated it is to define and measure teacher effectiveness, particularly in mainland China (Ko & Sammons, 2013; Muijs, 2006). In consideration of the use of the research findings, it would be very interesting to explore, for example, whether teachers in some rural schools could possibly exhibit more effective teaching practices than those in urban schools, since it is predictable that schools in urban areas will have teachers who exhibit more effective teaching practices than schools in rural areas.

3.2.2 Empirical procedures

The empirical procedures regarding research design, data sampling, data collection and data analysis will be identified below (see the elements outlined in table 2 for details).

3.2.2.1 Strengths

All the stages of the study in relation to research design, data collection and data analysis procedures have been examined thoughtfully. The procedures have been reported in sufficient detail to enable the research to be replicated.

Also, the research design is appropriate for the research questions. The authors have proposed research hypotheses and questions to examine whether more effective schools will have teachers who exhibit more effective teaching practices and whether schools from urban areas will have teachers who exhibit more effective teaching practices than those from rural areas. The authors have employed a parallel mixed-methods (QUAN + qual) approach (Onwuegbuzie & Leech, 2004, 2005, 2007) to test their hypotheses appropriately.

Furthermore, the data collection instruments and procedures for the quantitative part are adequate and appropriate for the research questions. The authors have presented relatively strong rationales of adopting the quantitative instruments, which have been piloted in other studies for improving reliability and validity (Teddle & Liu, 2008).

3.2.2.2 Weaknesses

The data collection instruments and procedures for the qualitative data are not adequate enough. It is difficult to get more in-depth and high quality data if the authors have only collected it via informal notes or school-level observations.

Crucially, the sample is not appropriate enough. The criteria of selecting effective and ineffective schools need to be further considered. The effective and ineffective schools have been selected based on test scores and expert opinions. Schools identified as effective, if looked at by using multilevel modelling techniques over time, might not be necessarily effective (Thomas et al., 2007). Also, the choices of 8 urban and 4 rural schools underrepresent rural schools, and no secondary schools have been sampled for classroom observations.

Also, the qualitative data analysis procedures are not adequate and appropriate for the research questions. The authors have only presented how the quantitative data has been analysed. The data analysis techniques for the qualitative data have been omitted.

3.2.3 The quality of the data

The quality of the data will be evaluated in terms of the reliability, validity and reactivity (see table 2 above for details). The strengths and weaknesses will be identified.

3.2.3.1 Strengths

The authors have been transparent in presenting their data collection procedures. The reader can be clear about who participated in the collection of data, how they have worked collaboratively to develop consistent and standardised implementation procedures and when the data was collected. Specifically, the reliability of the quantitative data is high (Punch, 2009), as this study is an ISERP replication and the instruments have been pilot tested in other studies. Strong rationales of adopting the instruments have also been given. In addition, the authors have informed the participants about the study in simple and unthreatening terms, have taken steps to minimise the effects, and they have tried to improve the reactivity of the data (Punch, 2009).

3.2.3.2 Weaknesses

The validity of the quantitative data can be questioned due to the underrepresentation of rural schools. Also no secondary schools are sampled in the study and the geographical differences are not considered thoughtfully, either between developed and underdeveloped areas or amongst urban, suburban and rural areas. In addition, the dependability (validity) of the qualitative data (Punch, 2009) could be critiqued since no specific data analysis techniques have been presented.

3.2.4 The findings and conclusions reached in the research

The key to the evaluation of the findings and conclusions is whether the research questions have been answered. The generalisability and applicability need also be considered. The strengths and weaknesses will be identified.

3.2.4.1 Strengths

Overall, the research questions have been answered. The findings of the study have successfully tested the research hypothesis proposed and have confirmed the previous findings regarding the educational differences between urban and rural areas. Teaching

practices that are ‘indigenous to China’ have also been identified to answer the research question, though it would be better to employ the same research methodology to conduct the research in a larger sample size, thereby improving its generalisability and applicability.

3.2.4.2 Weaknesses

The findings of the qualitative data have tended to be quite descriptive. The authors have tried to describe what they have observed in classrooms and report them directly in the research text, instead of explaining the reasons to facilitate a deeper understanding of those effective teaching practices that may be different from the international literature. Therefore, the qualitative findings are not sufficiently reflective. Also, there are no comparisons between the quantitative and qualitative data since this is a parallel mixed-methods study.

4. Improvements that could be made to the research

First, the authors should make it clearer what position the research is coming from, in order to enhance the coherence between the research aim and research design. A mixed-methods study would still be suggested to advance our understanding of the educational effectiveness processes, but it is crucial to balance the weight of the quantitative and qualitative parts accordingly. As long as the study has been set up for our further understanding, the qualitative part of the study needs to be emphasised in order to provide more in-depth interpretations. For example, unstructured, semi-structured or structured focus group interviews could be conducted to get in-depth and high quality data. Even ethnographic or narrative inquiry could be utilised to identify different teaching practices in mainland China.

Second, it is important to improve the sampling strategies. The criteria used for selecting effective and ineffective schools need to be made clearer. More rural schools, especially schools in the suburban areas, need to be sampled. Secondary schools need to be sampled in the study as well, making the research findings more persuasive and more generalisable. Also, the qualitative data analysis procedures should be more adequate and appropriate for the research questions and the data analysis techniques for the qualitative data should be included.

Third, based on the improvements mentioned above, the authors should consider further how to improve the quality of the data more appropriately in terms of its reliability (dependability), validity and reactivity. Finally, the findings of the qualitative data should be explained and reported in more depth and could be discussed and compared with the quantitative findings.

5. Conclusion

Based on the arguments above, it is worth addressing here how importantly the philosophical approach can impact on the coherence between research design and research methodology. The researcher has to be very clear about what they wish to explore and how they would like to undertake the research by employing the most appropriate philosophical and

methodological approaches.

Combining quantitative and qualitative methods would largely allow the researcher to explore the 'breadth' and 'depth' of the research findings, thereby providing the most suitable answer for research questions. Quantitative methods are helpful for the researcher to look at the breadth of the research findings, whilst qualitative methods look at the depth of the findings. However, the main concern is how to balance the weight of quantitative and qualitative methods when designing the research, based on what paradigms. Also no matter what specific mixed methods design would be, the researcher needs sufficient time to collect and analyse data. So will the time allow the researcher to do that in reality? Nevertheless, it is crucial for the researcher to know when to use a mixed methods approach, how to fit it in the research design and, more importantly, how to balance the weight of both quantitative and qualitative methods.

Acknowledgement

The research is financed by the China Scholarship Council. No. 201508510093. Special thanks go to Dr Sheila Trahar, Professor Sally Thomas and Dr Guoxing Yu at the University of Bristol.

References

- Campbell, R.J., Kyriakides, L., Muijs, D., & Robinson, W. (2003). Differential Teacher Effectiveness: Towards a Model for Research and Teacher Appraisal. *Oxford Review of Education*, 29(3), 347-362. <http://dx.doi.org/10.1080/03054980307440>
- Crossley, M. (2010). Context Matters in Educational Research and International Development: Learning from the Small States Experience. *Prospects*, 40(4), 421-429.
- Crossley, M. (2014). Global league tables, big data and the international transfer of educational research modalities. *Comparative Education*, 50(1), 15-26. <http://dx.doi.org/10.1080/03050068.2013.871438>
- Crotty, M. (1998). *The Foundations of Social Research: meaning and perspective in the research process*. London: Sage.
- Guba, E.G., & Lincoln, Y.S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In Denzin, N.K., & Lincoln, Y.S. (eds.), *The Sage Handbook of Qualitative Research, 3rd Edition*. Thousand Oaks, CA: Sage. pp. 191-215.
- Johnson, R.B., & Onwuegbuzie, A.J. (2004). Mixed Methods Research: A Research Paradigm. *Educational Researcher*, 33(7), 14-26.
- Johnson, R.B., Onwuegbuzie, A.J., & Turner, L.A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), 112-133.

- Ko, J., & Sammons, P. (2013). *Effective Teaching: A Review of Research and Evidence*. CfBT Education Trust.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Muijs, D. (2006). Measuring Teacher Effectiveness: Some Methodological Reflections. *Educational Research and Evaluation*, 12(1), 53-74. <http://dx.doi.org/10.1080/13803610500392236>
- Onwuegbuzie, A.J., & Leech, N.L. (2004). Enhancing the Interpretation of “Significant” Findings: The Role of Mixed Methods Research. *The Qualitative Report*, 9(4), 770-792.
- Onwuegbuzie, A.J., & Leech, N.L. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. *International Journal of Social Research Methodology*, 8(5), 375-387.
- Onwuegbuzie, A.J., Witcher, A.E., Collins, K.M., Filer, J.D., Wiedmaier, C.D., & Moore, C.W. (2007). Students' Perceptions of Characteristics of Effective College Teachers: A Validity Study of a Teaching Evaluation Form Using a Mixed-Methods Analysis. *American Educational Research Journal*, 44(1), 113-160.
- Peng, W., McNess, E., Thomas, S., Wu, X., Zhang, C., Li, J., & Tian, H. (2014). Emerging Perceptions of Teacher Quality and Teacher Development in China. *International Journal of Educational Development*, 34, 77-89. <http://dx.doi.org/10.1016/j.ijedudev.2013.04.005>
- Philips, D.C. (2004). Two Decades After: “After The Wake: Postpositivistic Educational Thought”. *Science & Education*, 13, 67–84.
- Punch, K., & Onacea, A.E. (2014). *Introduction to Research Methods in Education*. Los Angeles, Sage (Chapters 1 and 2).
- Punch, K.F. (2009). *Introduction to Research Methods in Education*. London: Sage.
- REF (2012). Panel criteria and working methods. Retrieved 18 November, 2015 from <http://www.ref.ac.uk/pubs/2012-01/>
- Reynold, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: International perspectives on school effectiveness*. London: Routledge/Falmer.
- Reynolds, D. (2006). World Class Schools: Some methodological and substantive findings and implications of the International School Effectiveness Research Project (ISERP). *Educational Research and Evaluation*, 12(6), 535-560. <http://dx.doi.org/10.1080/13803610600874026>
- Sun, H. (2011). wo guo jiaoyu xiaoneng yanjiu xianzhuang, wenti yu fazhan qushi (The Current Status Quo, Problems and Developmental Trend of Educational Effectiveness Research in China). *Journal of Shenyang Normal University (Social Science Edition)*, 35(167), 39-43.

- Sun, M. (2007). *jiaoyu xiaoneng lun (Educational Effectiveness Research)*. Beijing: Renmin Education Press.
- Teddlie, C., & Liu, S. (2008). Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People's Republic of China. *School Effectiveness and School Improvement*, 19(4), 387-407. <http://dx.doi.org/10.1080/09243450802535182>
- Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects*. New York: Teachers College Press.
- Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology With Examples. *Journal of Mixed Methods Research*, 1(1), 77-100.
- Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The International System for Teacher Observation and Feedback: Evolution of an International Study of Teacher Effectiveness Constructs. *Educational Research and Evaluation*, 12(6), 561-582. <http://dx.doi.org/10.1080/13803610600874067>
- Thomas, S., Peng, W., & Gray, J. (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33(3), 261-295. <http://dx.doi.org/10.1080/03054980701366116>
- Thomas, S.M., & Peng, W. (2011). Methods to evaluate educational quality and improvement in China. In Ryan, J. (ed.), *Education Reform in China: Changing concepts, contexts and practices*. Routledge: Contemporary China Series.
- Thomas, S.M., Kyriakides, L., & Tony, T. (2015). Educational Effectiveness Research in New and Emerging Contexts. In Chapman, C., Muijs, D., Reynolds, D., Sammons, P., & Teddlie, C. (eds.), *The International Handbook of Educational Effectiveness: Research, Policy and Practice*. Routledge: London.
- Thomas, S.M., Salim, M.M., Munoz-Chereau, B., & Peng, W. (2012). Education Quality, Effectiveness and Evaluation: Perspectives from China, South America and Africa. In Chapman, C., Armstrong, P., Harris, A., Muijs, Reynolds, D., & Sammons, P. (eds.), *School Effectiveness and Improvement Research, Policy and Practice: Challenging the orthodoxy?* Routledge: London and New York.
- Usher, R. (1996). A critique of the neglected epistemological assumptions of educational research. In Scott, D., & Usher, R.(eds.), *Understanding Educational Research*. Routledge: London.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).